

## When Is a Prognostic Factor Useful?: A Guide for the Perplexed

By M.N. Levine, G.P. Browman, M. Gent, R. Roberts, and M. Goodyear

Traditionally, a number of variables have been used to predict outcome in patients with early-stage breast cancer. These tests are simple to perform and relatively inexpensive. Recently, a number of new factors, eg, tumor proliferative index, nuclear DNA content, and amplification or overexpression of growth-promoting genes or oncogenes have been identified as potential predictors of outcome in patients with breast cancer. There is now increasing pressure to introduce such tests into routine clinical practice. How does a clinical practitioner identify which test, or group of tests, best predicts adverse outcome and whether any more clinically useful information is provided than with the use of more traditional factors alone? The aim of a prognostic test in breast cancer is to predict which patients are destined to develop a recurrence of cancer and those who are not. The prognostic usefulness of a test can be expressed in terms of relative risk (RR), which is the ratio of the risk of breast cancer recurrence in patients who test positive to the risk in those

who test negative. Methodologic guidelines that should be satisfied by a study evaluating the predictive ability of a test include the following: (1) Was an inception cohort assembled? (2) Was the referral pattern described? (3) Were laboratory and clinical outcomes assessed in a blinded fashion? (4) Was complete follow-up achieved? (5) Was adjustment for extraneous prognostic factors carried out? (6) Were appropriate statistical methods used? An approach is suggested to help the clinician choose the test, or combination of tests, likely to discriminate between "high-" and "low-risk" patients in his/her own practice. The decision regarding what particular threshold value (risk) defined by a prognostic test (or series of tests) warrants adjuvant therapy for an individual patient is a complex one but should be based on a clear presentation of the risks and benefits to the patient.

*J Clin Oncol* 9:348-356. © 1991 by American Society of Clinical Oncology.

**T**RADITIONALLY, a number of baseline characteristics, eg, number of involved axillary nodes, tumor size, hormone receptor status, and histology, have been used to predict outcome in patients with early-stage operable breast cancer.<sup>1</sup> In recent years, based on experimental studies in the laboratory at the cellular and molecular levels, there have been a number of clinical studies examining the prognostic ability of newer tests. Interest in such tests has been stimulated by clinical trials of adjuvant systemic therapy in patients with node-positive, and more recently, node-negative breast cancer.<sup>2-4</sup> The prediction of clinical outcome is particularly important for the latter group of patients in whom treatment could

potentially be tailored to the patient's risk of recurrent breast cancer (although high failure rate may not be synonymous with high responsiveness to adjuvant therapy). The recent National Cancer Institute (NCI) Clinical Alert has been interpreted by many oncologists as recommending adjuvant systemic chemotherapy for all women with node-negative breast cancer.<sup>5,6</sup> Considerable controversy has ensued and the arguments are well documented.<sup>5-7</sup> One approach that seems to be emerging is to reserve adjuvant therapy for those node-negative patients who can be identified as being at high risk for recurrence and death. Newer prognostic tests have been, and are being, evaluated in the research setting, but there is now increasing pressure on the oncologist to introduce them into routine clinical practice.

The clinical oncologist and laboratory physician are faced with an array of new prognostic factors including among others, DNA ploidy, S phase fraction labeling index, neu oncogene expression and amplification, epidermal growth factor receptor expression, and activity of various proteases, eg, cathepsin D.<sup>8-16</sup> How does a clinician identify which test, or group of tests, best predicts adverse outcome and whether any more clinically useful information is provided than by the use of more

---

*From the Departments of Medicine, and Clinical Epidemiology and Biostatistics, McMaster University; Ontario Cancer Treatment and Research Foundation Hamilton Regional Centre; and the Hamilton Civic Hospitals Research Centre, Hamilton, Ontario, Canada.*

*Submitted May 21, 1990; accepted July 20, 1990.*

*Dr Levine is a Scientist of the Medical Research Council of Canada.*

*Address reprint requests to M.N. Levine, MD, MSc, OCF-Hamilton Centre, 711 Concession St, Hamilton, Ontario, Canada L8V 1C3.*

*© 1991 by American Society of Clinical Oncology.  
0732-183X/91/0902-0005\$3.00/0*

traditional factors alone? Such questions are clearly clinically relevant but are also important because of increasing health care costs.

This report focuses on methodologic issues that should be considered when identifying new prognosticators in early-stage breast cancer. Our objective is to examine some of the important methodologic issues related to studies of predictive tests in node-negative breast cancer in order to help clinical oncologists decide whether a test is likely to be useful in the management of their own patients. In addition, it is hoped that investigators will keep such issues in mind when publishing studies on new predictive tests in breast cancer. We will not attempt to judge which tests, old or new, are best for predicting outcome. Although the ideas contained in this discussion may be of general interest, their specific application assumes that there is a sizeable population of medical oncologists who agree with the strategy of reserving adjuvant chemotherapy for "high-risk" patients with stage I breast cancer.<sup>7</sup> It is also assumed that recommendations regarding treatment in individual patients occurs within the context of informed and open disclosure of information to patients in the decision-making process.

**METHODOLOGIC PRINCIPLES**

*A Test Used for Prediction*

Tests are used in clinical medicine for a number of reasons: (1) to determine whether a patient has a disease of interest (diagnosis), (2) to determine whether a patient will develop a certain condition (prediction), and (3) to follow the clinical course of a disease, eg, response to therapy. The characteristics of diagnostic tests are described by their sensitivity, specificity, and positive and negative predictive values.<sup>17</sup> These are easily understood by referring to a 2 × 2 table in which the test result is displayed according to the condition under study (Table 1).

The aim of a prognostic test in breast cancer is to predict for an individual patient the likelihood

of recurrent cancer or death. Studies designed to evaluate such tests generally attempt to show systematic differences in subsequent outcome between groups defined by the test result. In applying these data, the clinician must determine for an individual the test group to which she belongs. While the concepts of sensitivity, specificity, predictive values, true-positive and false-positive are useful in the diagnostic situation in which a patient either does or does not have the disease of interest, they are slightly less satisfactory in the time-dependent prognostic situation. If, for example, one is predicting disease recurrence in the 5-year period following initial surgery and a high-risk patient does not develop recurrent disease, is it fair to say this is a "false-positive"? The fact that a patient has not developed the disease in the first 5 years after surgery, does not mean that she is no longer to be considered high risk. She may develop it in the sixth or seventh year. In this particular situation the inherent prognostic usefulness of a test is usually expressed in terms of the relative risk (RR). For a test that is categorized as either positive or negative, this is the ratio of the risk of the bad outcome in patients who test positive to the risk in those who test negative (Table 2). A RR of 2 for recurrence means that the patient is twice as likely to have a recurrence if the test result were positive than if the test were negative.

Since predictive tests deal with future events, their validation requires that the time of follow-up for the study population, which will contain patients of varying risk, be standardized in some way. This can be done by studying the cumulative risk in individuals in the population who have had a minimum follow-up, eg, 5 years, if we were interested in the 5-year survival rate for a given test result. A more efficient method, which allows us to use the experience of individuals who have been followed for shorter periods, is to use survival curves. This has the additional advantage of providing information about failure patterns throughout the entire follow-up period. Thus, most published reports of predictive tests in breast cancer patients

**Table 1. Characteristics of a Diagnostic Test**

Test	Breast Cancer	
	Yes	No
Positive	True-positive, a	False-positive, b
Negative	False-negative, c	True-negative, d

NOTE.  $a/a + c$  = sensitivity;  $d/b + d$  = specificity;  $a/a + b$  = positive predictive value;  $d/c + d$  = negative predictive value.

**Table 2. Characteristics of a Predictive Test**

Test	Recurrent Breast Cancer	
	Yes	No
Positive	a	b
Negative	c	d

NOTE.  $RR = a/a + b/c/c + d$ .

compare test subgroups in terms of their survival curves. The optimal scenario for a prognostic test is depicted in Fig 1. A more likely situation, however, is shown in Fig 2 in which "the best" available prognostic test divides patients into "high-" and "low-risk" subgroups that have vastly different subsequent risks but still with a degree of uncertainty since not all the high-risk group patients will recur nor all the low-risk group patients escape the bad outcome. If the test results were used to decide treatment, more patients would receive an appropriate therapy, but not without the occurrence of some over- or under-treatment.

#### Performance of Tests in the Laboratory

When reading about a new prognostic test in the literature, the oncologist should attempt to judge whether the subsequent application of the test is going to be technically possible in his or her own setting. The issues related to the technical aspects of laboratory tests have been reviewed by Callahan and Campbell.<sup>18</sup>

Having determined that a test is valid, based on an adequate study design (see below), technically feasible, and affordable, the oncologist must then obtain information about its reproducibility, within and between laboratories, and its generalizability to the population to which it will be applied. For example, over the years, quality assurance mechanisms have been developed, including the establishment of reference centers, to maintain reasonable interlaboratory agreement for hormone receptor assays.

The alert oncologist should consider the nature of the data that are generated by a predictive test before applying it in practice, since this may affect

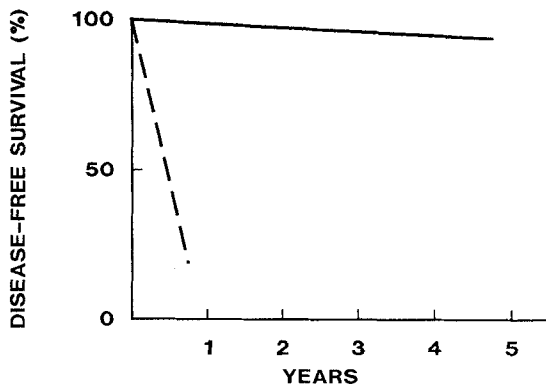


Fig 1. The ideal prognostic test. (—) Low risk, (----) high risk.

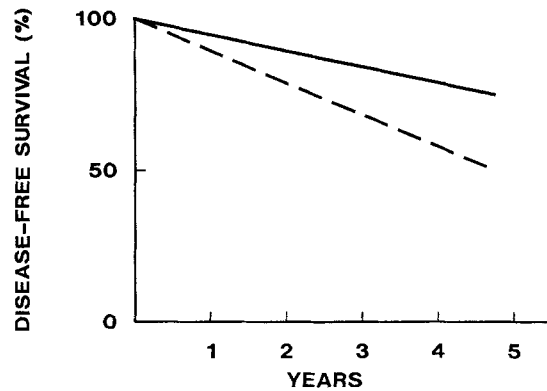


Fig 2. The prognostic test likely to be encountered in clinical practice. (—) Low risk, (----) high risk.

how the results are interpreted. Certain variables, such as menopausal status, are categorical. For histologic grade, the categories (I, II, and III) are also well defined but are ordered in a meaningful way with respect to prognosis.<sup>18-20</sup> For DNA ploidy the test is either normal (diploid) or abnormal (aneuploid).<sup>8,9</sup> Similarly, when using immunohistochemical staining to determine overexpression of the neu oncogene, overexpression is reported as either present or absent.<sup>14</sup> Variables such as age, tumor size, and hormone receptor status are continuous, since each patient observation theoretically falls somewhere along a continuum. However, in many instances, often for convenience, such variables are arbitrarily classified into two categories and then are considered dichotomous. In the breast cancer literature, age is often divided into younger than 50 years and 50 years or older, tumor size into less than 2 cm and  $\geq 2$  cm, and hormone receptor status into less than 10 fmol and  $\geq 10$  fmol.

Proliferative capacity as measured by DNA S phase fraction is a continuous variable. However, in published reports, it is often expressed as categorical. Some investigators choose the median of the distribution to divide the population into two equal-sized groups.<sup>8</sup> However, this may not be the most appropriate method for categorizing patients. (The shape of the relationship between risk and test result may result in the cutoff with the best discriminating value not being at the median.) Another method that has been used to choose the optimal cutoff for S phase is to choose the value that yields the most appropriate balance between positive and negative predictive values.<sup>8,17</sup> The

choice of the best cutoff is not an easy one to make, but one strategy that can reassure the reader as to the validity of a certain cutoff would be the strategy of determining a cutoff to discriminate groups in one cohort of breast cancer patients and then validating this cutoff in a second group of breast cancer patients. This strategy was used in the report by Sigurdsson et al.<sup>9</sup> Finally, the cutoff value obtained for a given population may apply only to the population from which it is derived and may not be generalizable to other populations. Thus, within the same institution, cutoffs identified in a population of patients with stage II breast cancer may not be appropriate to identify different risk categories for patients with stage I disease and, similarly, the generalizability of cutoff values between institutions is clearly problematic. Different optimal cutoff values for S phase in stage I breast cancer have been reported.<sup>8,9,21-23</sup>

*Guidelines for Assessing Studies on Natural History*

Studies of prognostic tests that follow cohorts of patients forward in time are subject to biases that can be avoided with proper attention to experimental design. When evaluating published studies on prognostic tests in breast cancer, the oncologist should be aware of certain pitfalls in the design, execution, and analysis that may threaten the validity of the study.<sup>17,24</sup> Guidelines for critically appraising published studies on prognostic tests in breast cancer are presented in Table 3.

An inception cohort of patients should be assembled, that is, patients should be identified at an early and uniform point (inception) in the course of their disease. Breast cancer patients should be identified postmastectomy and followed prospectively, so that both those who develop recurrence of breast cancer and those who remain disease-free are included in the study. The type of surgery, period of patient recruitment, and nature of adjuvant therapy should be described. A clearly defined inception cohort can minimize the opportu-

nity for bias by avoiding the selection of patients at a particularly high or low risk of recurrent disease.<sup>25</sup> In the study by Slamon et al on the neu oncogene, tissue samples were obtained from breast cancer patients who were part of an ongoing study at the University of Texas, San Antonio.<sup>11</sup> There was, however, no description of the nature of this study, the time period over which patients were recruited, and the type of surgery these patients had.<sup>11</sup> In contrast, in the study reported by Sigurdsson et al<sup>9</sup> on S phase and ploidy, a clear description of the patients studied was provided. They were consecutive patients who underwent surgery for node-negative breast cancer between 1982 and 1986. The baseline disease characteristics of these patients and type of adjuvant radiation they received were also described in detail.

The referral pattern of patients entered into the study should be described. This can help to determine whether the study results are generalizable to patients in one's own practice. For example, patients referred to a tertiary cancer center may have more severe disease than the average patient (referral bias).<sup>25</sup> Presumably, this can be captured by examining the baseline characteristics of the patients. This particular issue is illustrated by the results of a recent randomized trial of adjuvant chemotherapy in stage I breast cancer reported by Bonadonna et al.<sup>26</sup> The chemotherapy substantially reduced the risk of recurrent breast cancer, but the trial has been criticized because the failure rate of approximately 50% in the control group was felt by some to be much higher than that anticipated in the "average" patient with stage I breast cancer.

In studies of prognostic tests, the assessment of important clinical and laboratory outcomes should be performed in a blinded fashion. The performance and interpretation of laboratory test results by an individual who could be aware of the clinical outcomes should alert the reader to the possibility of unconscious bias affecting the reporting of laboratory results. Although the results themselves may not have been tampered with, decisions about inclusion and exclusion of some test results may have been influenced (eg, the adequacy of a pathologic preparation). Similarly, if a chart review is performed to obtain information on clinical events, it should be done by individuals unaware of the prognostic test result. In nine recent publications on new prognostic factors in operable breast

**Table 3. Criteria for Critically Appraising Studies of Prognostic Tests in Early-Stage Breast Cancer**

Was an inception cohort assembled?
Was the referral pattern described?
Were laboratory and clinical outcomes assessed in a blinded fashion?
Was complete follow-up achieved?
Was adjustment for extraneous prognostic factors carried out?
Were appropriate statistical methods used?

cancer,<sup>8-16</sup> only two stated that the laboratory tests were performed by individuals unaware of clinical outcome.<sup>11,14</sup>

Outcome results should be available for all patients who entered the inception cohort. Failure to achieve complete outcome ascertainment may lead to a biased assessment of the predictive ability of a test, since reasons for losses to follow-up may not be distributed evenly between high- and low-risk subgroups. For example, if survival results were not available for a group of patients in the cohort, which was potentially at high risk for an adverse event, then failure to include their outcome information could bias the predictive ability of the test. The degree of follow-up for all patients entered into the study should be clearly stated.

Other factors that are potentially associated with the disease in question and that could affect the prognosis should be clearly described and adjusted for in the analysis. For example, in a study evaluating a new predictive test in premenopausal women with node-negative breast cancer, adjuvant chemotherapy could influence outcome. If more women with a risk factor that was thought to identify a high-risk group received adjuvant chemotherapy, then this could bias the results. In eight recent studies examining the predictive ability of prognostic factors in early-stage breast cancer, many of the patients received adjuvant chemotherapy.<sup>8,9,11-16</sup> However, in only one of these studies was there an attempt in the analysis to adjust for the effect of the chemotherapy on outcome.<sup>9</sup>

Appropriate statistical methods should be used in the study. The results should be analyzed and presented in such a way that clinically relevant predictions for outcome can be made. (This will be discussed at greater length later.) In addition, sufficiently large numbers of patients should be included to provide the reader with confidence that the observed results are correct.

When reading published studies on prognostic tests in breast cancer, the oncologist should consider the nature of the study design. Prospective studies such as the randomized control trial and cohort analytic study (concurrent control, but not randomized) are more likely to satisfy the methodologic guidelines discussed above than retrospective ones because better control over such important components as selection of the study

population, consistency in the application of the test, and appropriate assessment of outcome criteria can be maintained. Consequently, there is less opportunity for bias. A type of study design that is presently often used to identify prognosticators in breast cancer is what we will call "a historical prospective cohort study." In this design, a group of patients who have undergone surgery for breast cancer in the past is identified and then followed forward to the present time for recurrence or survival (often they are or were in a randomized trial). The predictive test is performed on pathologic tissue that has been retrieved from the original surgery, and the result is related to patient outcome. This design is obviously more efficient than initiating a cohort and following the patients for a number of years until there are sufficient events. However, this design depends on the ability to retrieve archival material on all patients. The reader should verify that the patients studied are representative of the original cohort from which they were selected to ensure that bias is minimized. In a recent report on erbB-2 overexpression, pathologic material was obtained from 292 patients from the National Surgical Adjuvant Breast and Bowel Project protocol B-06, and methods of selection were not stated.<sup>14</sup> There were, however, over 1,000 patients originally recruited in this trial, and an obvious issue is whether the study sample is truly representative of the entire population.

In a case-control study (retrospective), breast cancer patients with an adverse outcome, eg, recurrence or death, are compared with breast cancer patients who have not experienced such an event in terms of the predictive test. This particular study design is subject to greater opportunities for the introduction of bias compared with other study designs.<sup>25</sup>

#### *Statistical Analysis*

When reading published studies on prognostic tests in breast cancer, the oncologist often is faced with a variety of statistical analyses that can appear quite complicated. An in-depth discussion of analyses for prognostic tests is beyond the scope of this report, but certain salient features will be highlighted.

A particular pattern of statistical analysis has emerged in published studies of prognostic tests in breast cancer. A simple way to identify prognostic

variables is to examine individually the relationship of each variable to the length of survival or remission. For each variable, patients are grouped according to different cutoffs of the variable, yielding subgroups, for example, less than three nodes versus three or more nodes. Actuarial survival curves for the various subgroups defined by the variable of interest are presented and then compared to determine risk categories using a statistical test, eg, the log-rank test.<sup>27</sup> A second variable is selected, broken down into defined subgroups, and the survival experience of these groups is again compared to determine risk categories. This process may then be repeated for any number of variables. Such an approach to analysis is termed "multiple univariate comparisons" and can be problematic for two reasons. First, the more comparisons that are tested statistically, the more likely one is to find a "statistically significant" relationship by chance alone.<sup>28</sup> This may account for some of the inconsistencies observed between studies. The reader is more likely to be reassured that a particular variable defines a high-risk group if that variable consistently predicts outcome in several studies. The second problem is how to handle the predictive value of combinations of tests since individual tests, such as menopausal status, histologic grade, and hormone receptor status may be highly correlated. For example, patients who are premenopausal are more likely to be estrogen receptor (ER)-negative than postmenopausal women, and ER-negative tumors are more likely to have a poor histologic grade. In an attempt to address this problem, published studies will typically compare the survival experience of patient subgroups defined by a combination of tests. To illustrate, if menopausal status can be divided into pre- and postmenopausal, and ER into positive and negative, then there are four possible subgroups that can be characterized. The subsequent risk in each group can then be summarized in a survival curve or the proportion of subjects experiencing the bad outcome in a defined time period, ie, the survival of premenopausal ER-negative patients is compared with that of premenopausal ER-positive patients. However, this process reduces the number of patients in the subgroups of interest thereby reducing the power to detect a clinically important risk difference between the groups, a type II error.<sup>29</sup>

Since stratification by multiple prognostic fac-

tors will inevitably run out of data, analysts usually resort to a modeling approach. The model allows for the investigation of the relative importance of a number of correlated prognostic factors but requires one to make some assumptions about how two or more factors exert their influence together. The most commonly used approach is the Cox proportional hazards model.<sup>30,31</sup> The results of such an analysis provide information on whether a particular test contributes to prognosis independently, ie, when the contribution of other tests have been accounted (adjusted) for. If the *P* value associated with a particular factor in the model is significant, then the factor provides additional independent information on prognosis. As the name implies, the model's basic assumption is that the risk for a patient with a positive test will be a constant multiple of the equivalent risk for a negative-test patient over time. The risk multiplier produced by a positive test is simply the RR mentioned earlier and can be derived by a simple calculation.<sup>31</sup>

When modeling the combined influences of two tests it is natural to assume as a starting point that the two individual RRs multiply together if both tests are positive. Thus, if test 1 has an RR of 1.5 and test 2 one of 2, a patient with both tests positive would be at  $1.5 \times 2$  equals three times the risk compared with a patient with both tests negative. The fact that we are essentially implying by this multiplicative model that the two tests convey independent influences on risk does not mean that the test results themselves are uncorrelated.

The Cox regression model has been used in a number of recent publications on prognosticators in early-stage breast cancer.<sup>8,9,11-16</sup> In some of the studies, however, RRs were not presented.<sup>11-15</sup> In the report by Sigurdsson et al about patients with node-negative breast cancer, a Cox model was performed and examined for the effect of the following factors on outcome, S phase, tumor size, progesterone receptor (PR), ER, ploidy, and age.<sup>9</sup> Only S phase and age had a statistically significant influence on survival. The RR associated with S phase fraction greater than 12% was 2.1 and that with age less than 50 years was 0.4. This report also provides an example of how information from a Cox model was used to develop a predictive index based on the number of risk factors present. The survival of the subgroup of patients who had all

factors at high risk was calculated and compared with the survival of the subgroup with all high-risk factors absent and to the subgroup with only some high-risk factors present. The validity of such an index depends on the precision of the estimate of survival for the subgroup, ie, if there are very few patients in the subgroup, the "true" survival of the subgroup may fall within a wide range.

#### AN APPROACH TO INTERPRETATION OF PREDICTIVE MODELS

A number of variables that have been used to predict prognosis in breast cancer are simple to perform, have established interlaboratory reliability, and are relatively inexpensive. These include axillary nodal status, patient age, tumor size, ER status, PR status, and histologic grade (although there are some who may dispute the merits of this last test<sup>20</sup>). There are now a number of ongoing clinical trials in which not only are these routine tests performed, but tests such as neu oncogene, S phase, and ploidy are being performed prospectively on all patients. One can anticipate in the near future that results of analyses on the prognostic ability of all these tests, both old and new, will be published. How can it be determined whether a new prognostic test provides additional predictive information to the more traditional ones? The following approach is suggested when considering this question.

First, the Cox regression analysis is performed, which includes all of the above routine variables to determine which predicts independently for outcome. The results presented should include not only a *P* value, but also an RR and the number of patients in each category. Consider the following hypothetical data for a group of 200 women with node-negative breast cancer in which each factor is expressed dichotomously: age (< 50, ≥ 50), *P* = .3, RR = 1.1; size (< 2, ≥ 2), *P* = .001, RR = 2.1; ER (< 10, ≥ 10), *P* = .23, RR = 0.85; PR (< 10, ≥ 10), *P* = .03, RR = 1.7; and grade (high, low), *P* = .3, RR = 0.95.

Only size and PR predict independently for breast cancer recurrence as shown by the *P* values. In order to derive an overall RR for the population, these two RRs can be multiplied, ie,  $2.1 \times 1.7 = 3.6$  (for this approach, we have made the assumption that there are no interactions). This means that the risk of recurrence for a patient with

a tumor ≥ 2 cm and a PR less than 10 is 3.6 times that for a patient whose tumor is less than 2 cm and PR ≥ 10. There were 50 such patients with both high-risk factors present. Although a RR provides more clinically relevant information than a coefficient or *P* value, there are those who still might have some difficulty with this concept. In such cases, the relapse-free survival of patients with less than 2 cm and PR ≥ 10 tumors can be presented and compared to that of patients with tumors ≥ 2 cm and PR less than 10.

To determine whether a new test provides additional predictive information over that provided by more traditional tests, the Cox analysis is now repeated with all of the above more traditional variables plus the new test. Suppose DNA ploidy is added to the above tests in our cohort of 200 women with node-negative breast cancer. Now, not only size and PR predict independently for recurrence (RRs of 1.8 and 1.7, respectively), but also aneuploidy is associated with a RR of 2.0 compared with diploid tumors. The overall RR is now 6.1 ( $1.8 \times 1.7 \times 2.0$ ) for the 45 patients who have all three factors predicting a high risk of recurrence compared with patients without any of these high risk factors.

Consider another cohort of 200 patients with node-negative breast cancer who have all the above tests including ploidy performed. In this situation, only size and ploidy predict independently for breast cancer recurrence, with RRs of 1.5 and 2.5, respectively. The overall RR is now 3.7 ( $1.5 \times 2.5$ ). Notice, however, that with ploidy now added to the model, no additional clinically useful predictive information is provided compared with using only the more traditional tests as presented earlier. In addition, ploidy by itself might "look good," but its additional information might be low.

Although this section has dealt with analysis and methods to present results in a clinically meaningful way, the final decision of whether adjuvant treatment should be used is based on a clear presentation of the risks and benefits to the patient.

#### MAKING THE CLINICAL DECISION

Recently, results from a number of clinical trials have demonstrated that adjuvant chemotherapy in patients with stage I breast cancer reduces the recurrence rate by approximately 30% within the

first 4 years after initial surgery.<sup>24</sup> However, in some studies, this has not yet been translated into a survival benefit. The issue of whether to use adjuvant chemotherapy in stage I breast cancer patients has sparked much discussion and controversy.<sup>6,7</sup> The availability of prognostic tests does not relieve the clinician from the burden of making clinical decisions. The oncologist in our scenario must make two kinds of decisions: (1) which prognostic test(s) should be measured, and given the prognostic information, (2) what treatment option should be recommended?

If all prognostic tests provided identical information, then only one would be needed, and the choice would depend on feasibility and cost. For example, if it were found that tumor grade provided all of the information required for making a decision, then it would be unnecessary to incur the costs and time associated with obtaining further information. Whether or not further investigation is warranted depends on the answer to the following question, "At what risk of recurrence am I prepared to recommend treatment with adjuvant therapy in an individual patient?"

If the answer to the above question is that treatment is justified for any risk greater than that of the general population, then all patients with stage I breast cancer should receive therapy and it would not be necessary to measure any predictive tests. The clinician who is not prepared to follow this approach must then make a judgment regarding the risk of breast cancer recurrence that would justify therapy. Such judgments are commonplace in clinical medicine where clinicians must weigh potential benefits of an effective therapeutic intervention against the cost of side effects and other risks.

In one of the examples presented above, the RR of recurrence for patients with tumors  $\geq 2$  cm and PR less than 10 is 3.6 compared with patients with

tumors less than 2 cm and PR  $\geq 10$ . Although a risk of 3.6 appears substantial, it must be related to the average risk of the node-negative population as a whole and the low-risk group. For example, if the cumulative recurrence rate for patients with tumors less than 2 cm and PR  $\geq 10$  was 2% (low risk), then the cumulative risk for the high-risk group would be of the magnitude of 7.2% ( $2 \times 3.6$ ). The issue now, despite an established risk reduction for recurrence of approximately 30% with chemotherapy, is whether a 7.2% risk of failure warrants adjuvant therapy (and even performing the tests). Some patients (and physicians) may feel that this risk is small and not worth undergoing the morbidity of therapy, while other patients (physicians) may feel that a 7.2% risk is still significant and, hence, worth undergoing therapy. On the other hand, the decision to treat would likely be different if the cumulative recurrence rate for the low-risk group was 10%, but increased to 36% ( $10 \times 3.6$ ) in the high-risk group. Thus, decisions about treatment are related to underlying risks, the relative increase in risk conferred by a particular prognostic factor, and the relative effectiveness of the available intervention. Similarly, how should physicians deal with patients who have a mixture of prognostic factors (ie, some good and some bad)? The clinician needs to judge the quality of the published information, categorize a particular patient into high, intermediate, or low risk (perhaps more weight should be placed on the worst prognostic factor), and then present this information to the patient.<sup>32</sup>

The decision regarding what particular threshold value defined by a prognostic test (or series of tests) warrants adjuvant therapy for an individual patient is a complex one, but clearly a well-informed patient should be part of the decision-making process.

## REFERENCES

1. Merkel DE, Osborne KC: Prognostic factors in breast cancer. *Hematol Oncol Clin North Am* 3:641-652, 1989
2. Fisher B, Redmond C, Dimitrov NV, et al: A randomized clinical trial evaluating sequential methotrexate and fluorouracil in the treatment of patients with node-negative breast cancer who have estrogen receptor negative tumors. *N Engl J Med* 320:473-478, 1989
3. Mansour EG, Gray R, Shatila AH, et al: Efficacy of adjuvant chemotherapy in high-risk node-negative breast cancer. *N Engl J Med* 320:485-490, 1989
4. The Ludwig Breast Cancer Study Group: Prolonged disease-free survival after one course of perioperative adjuvant chemotherapy for node-negative breast cancer. *N Engl J Med* 320:491-496, 1989
5. Glick JH: Adjuvant therapy for breast cancer. *J Natl Cancer Inst* 80:471-475, 1988
6. DeVita VT Jr: Breast cancer therapy: Exercising all our options. *N Engl J Med* 320:527-529, 1989
7. McGuire WL: Adjuvant therapy of node-negative breast cancer. *N Engl J Med* 320:525-527, 1989

8. Clark GM, Dressler LG, Owens MA, et al: Prediction of relapse or survival in patients with node-negative breast cancer by DNA flow cytometry. *N Engl J Med* 320:627-633, 1989
9. Sigurdsson H, Baldetorp B, Borg A, et al: Indicators of prognosis in node-negative breast cancer. *N Engl J Med* 322:1045-1053, 1990
10. Hery M, Giovanni J, Lalanne CM, et al: The DNA labelling index: A prognostic factor in node negative breast cancer. *Breast Cancer Res Treat* 9:207-211, 1987
11. Slamon DJ, Clark GM, Wong SG, et al: Human breast cancer: Correlation of relapse and survival with amplification of the HER-2/neu oncogene. *Science* 235:177-182, 1987
12. Van De Vijver MJ, Peterse JL, Mooi WJ, et al: Neu-protein overexpression in breast cancer: Association with comedo-type ductal carcinoma in situ and limited prognostic value in stage II breast cancer. *N Engl J Med* 319:1239-1245, 1988
13. Wright C, Angus B, Nicholson S, et al: Expression of c-erb B-2 oncoprotein: A prognostic indicator in human breast cancer. *Cancer Res* 49:2087-2090, 1989
14. Paik S, Hazan R, Fisher ER, et al: Pathologic findings from the National Surgical Adjuvant Breast and Bowel Project: Prognostic significance of erb B-2 protein overexpression in primary breast cancer. *J Clin Oncol* 8:103-112, 1990
15. Sainsbury JRC, Farndon JR, Needham GK, et al: Epidermal-growth-factor receptor status as predictor of early recurrence of and death from breast cancer. *Lancet* 1:1398-1402, 1987
16. Tandon AK, Clark GM, Chamnes GC, et al: Cathepsin D and prognosis in breast cancer. *N Engl J Med* 322:297-302, 1990
17. Sackett DL, Haynes RB, Tugwell P: *Clinical Epidemiology: A Basic Science for Clinical Medicine*. Boston, MA, Little Brown, 1985, pp 47-138
18. Callahan R, Campbell G: Mutations in human breast cancer: An overview. *J Natl Cancer Inst* 81:1780-1786, 1989
19. Fisher B, Redmond C, Fisher ER, et al: Relative worth of estrogen or progesterone receptor and pathologic characteristics of differentiation as indicators of prognosis in node-negative breast cancer patients. *J Clin Oncol* 6:1076-1087, 1988
20. McGuire WL: Estrogen receptor versus nuclear grade as prognostic factors in axillary node-negative breast cancer. *J Clin Oncol* 6:1071-1072, 1988
21. Silvestrini R, Daidone MG, Canova S, et al: Node-negative breast cancers: A test bench for prognostic markers. *Breast Cancer Res Treat* 10:86, 1987 (abstr)
22. Hery M, Giovanni J, Lalanne CM, et al: The DNA labelling index: A prognostic factor in node-negative breast cancer. *Breast Cancer Res Treat* 9:207-211, 1987
23. Meyer JS, Friedman E, McCrate MM, et al: Prediction of early course of breast carcinoma by thymidine labelling. *Cancer* 51:1879-1886, 1983
24. Sackett DL, Haynes RB, Tugwell P: *Clinical Epidemiology: A Basic Science for Clinical Medicine*. Boston, MA, Little Brown, 1985, pp 159-169
25. Sackett DL: Bias in analytic research. *J Chronic Dis* 32:51-63, 1979
26. Bonadonna G, Valagussa T, Zambetti M: Milan adjuvant trials for stage I-II breast cancer, in Salmon SW (ed): *Adjuvant Chemotherapy of Cancer V*. Philadelphia, PA, Grune & Stratton, 1987, pp 211-221
27. Mantel N: Evaluation of survival data and two new rank order statistics arising in its consideration. *Cancer Chemother Rep* 50:163-170, 1966
28. Armitage P, McPherson K, Rowe BC: Repeated significance tests on accumulating data. *J R Stat Soc* 132A:235-244, 1969
29. Freiman J, Chalmers TC, Smith H, et al: The importance of beta, the type II error and sample size in the design and interpretation of the randomized controlled trial. *N Engl J Med* 299:690-694, 1978
30. Cox DR: Regression models and life-tables. *J R Stat Soc (B)* 34:187-220, 1972
31. Tibshirani R: A plain man's guide to the proportional hazards model. *Clin Invest Med* 5:63-68, 1982
32. McGuire WL, Tandon AK, Allred DC, et al: How to use prognostic factors in axillary node negative breast cancer patients. *J Natl Cancer Inst* 82:1006-1015, 1990